

Question 1:

- (a) When we want to estimate the causal effect of a treatment on some outcome, in here is mental health court effect on recidivism, we need to make sure that there's no selection bias due to self-selection into treatment, which is $E[Y^0 | D=1] \neq E[Y^0 | D=0]$. This basically means that the potential outcome of recidivism of people went to and didn't go to mental health court should not be equal. Also, $E[Y_i^0 | X=c_0]$ and $E[Y_i^1 | X=c_0]$ are continuous in X at c_0 , which means at score 60 in the mental illness scale, potential outcomes on recidivism of the treated and untreated groups should be continuous.
- (b) I think this colleague is wrong, because the court sets up this mental illness scales is to help people with mental illness, which by definition is correlated with the potential outcome, so here has the selection bias problem because of the self-selection into treatment. So the way this colleague recommended won't work.
- (c) running variable should be the mental illness test score. The cutoff should be scores at 60 points. People who intentionally make their test scores become higher than 60 would be the counterfactual inmate for each defendant in mental health court.
- (d) The key identifying assumption here is that $E[Y_i^0 | X=c_0]$ and $E[Y_i^1 | X=c_0]$ are continuous in X at c_0 , which here means that test scores should be continuous for both mental health courts' and normal courts' potential outcome on recidivism.
- (e) we could run a McCrary density test to see if the density is continuous at the cutoff points, which is 60 score in mental test here. If there's discontinuity around the cutoff, then manipulation problem might be an issue here. One good way is to use visualization

techniques to test on the density of the test scores here. We could plot number of observations in each bin of scores, for example 5 points as a bin, this plot should allow us to investigate whether there's discontinuity or heaping in the distribution of the mental health test scores at a 60 point bin. If there's heaping or discontinuities in the density plot, this suggests that people can manipulate the running variable scores, which is the mental health score in this example, which violates the smoothness as well.

(f) A Balance test should has smoothness through the cutoff c_0 , which means the density through 60 points mental health score should has smoothness.

$$(g) \text{Recidivism} = \alpha + \beta_{01} \tilde{S}_i + \beta_{02} \tilde{S}_i^2 + \dots + \beta_{0p} \tilde{S}_i^p + \delta D_i + \beta_1^* D_i \tilde{S}_i + \beta_2^* D_i \tilde{S}_i^2 + \dots + \beta_p^* D_i \tilde{S}_i^p + \varepsilon_i$$

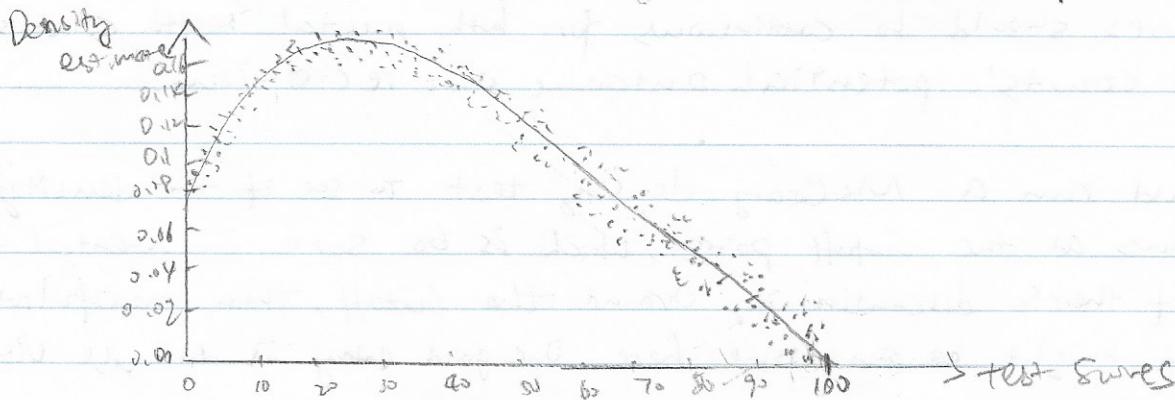
$$\text{where } \beta_{01}^* = \beta_{01} - \beta_{02}; \beta_2^* = \beta_{21} - \beta_{22}; \beta_p^* = \beta_{1p} - \beta_{0p}$$

② treatment effect at c_0 (score 60) is δ

③ $S_i = \text{Score}_i - 60$ (centered mental health score)

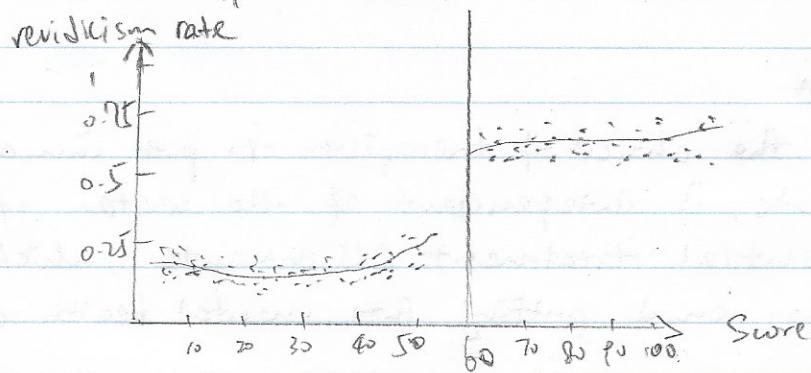
④ $D_i = \text{Assignment to the health court}$

(h) ① first, we need to have a density test figure, which is to plot the frequency of different test scores happened. The underlying hypothesis is that we need continuity in this plot, if we don't have continuity, they might have manipulation problems

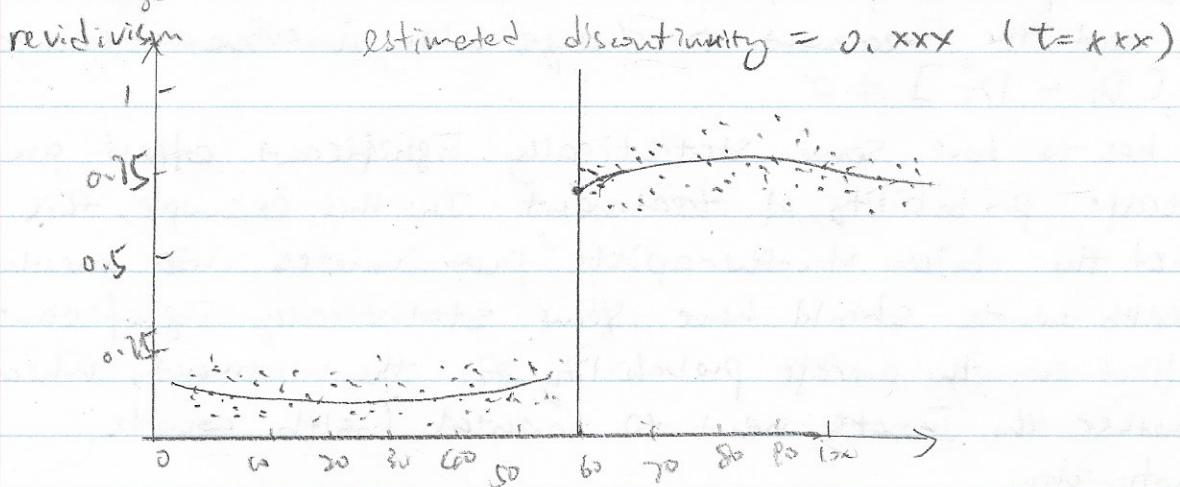


(2)

- ② second, after we plot the density plot and find out that there's no manipulation problem, we could plot an "outcome by running variables with small bins" plot. This is to check if we choose to use RDD is an optimistic way by take a closer look at the dependent variable. There should have a gap at the cutoff.



- ③ We could also do a "probability of treatment" plot to check the probability of discontinuity at cutoff is how large. In a fuzzy RDD, we want to see that the treatment variable jumps at the cutoff point; this could tell us whether we have a first stage.



Question 2:

(a) The choice of therapists to put inmates into mental health court

(b) ① Stable unit treatment value assumption (SUTVA)

Under this assumption, potential outcomes of reidivism for each inmates are unrelated to the treatment status of other individuals, which is whether other inmates went to the mental health court.

② Random assignment

The IV, which is the choice of therapists to put inmates into mental health courts, is independent of the vector of potential outcomes and potential treatment assignments, which here are the reidivism and putting into mental health court.

③ exclusion restriction

Any effect of Z on y must be via effect of Z on D .

This means that the only effect from the IV to potential outcome should come from the effect of IV on treatment.

In other words, $T(D_i, Z_i)$ is a function of D only.

④ Nonzero first Stage

We need the treatment to change when instrument changes.

$$E[D'_i - D_i] \neq 0$$

Z has to have some statistically significant effect on the average probability of treatment. In this example, this means that the choices of therapists put inmates into mental health courts should have some statistically significant effect on the average probability of the treatment, which is whether the inmate went to mental health courts.

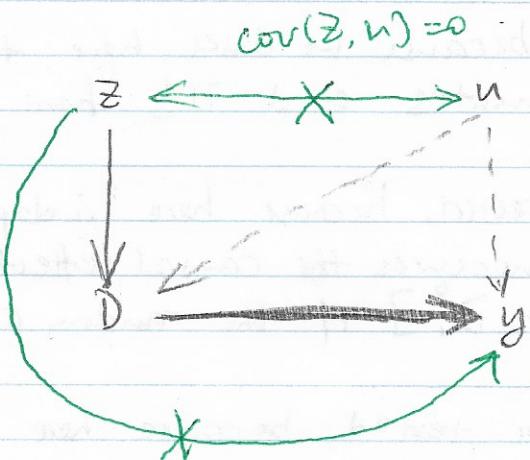
⑤ Monotonicity

This requires that the IV operate in the same direction on all individual units. In other words, while the instrument have no effect on some people, all those who are affected are affected in the same direction (positively or negatively, not both)

(3)

- (c) ① SUTVA could not be tested, because we don't have the potential recidivism of inmates and it's hard to verify whether $Z_i = Z_j$.
- ② Random Assignment could be tested, because here independence means that the first stage measures the causal effect of Z_i on D_i , which is $E[D_i^* - D_i]$ if the random assumption holds.
- ③ Exclusion restriction could be tested, because here we just want to make sure two things: ① $\text{Cov}(Z, u) = 0$; ② Z only affects y through its effect on D . It's possible to use placebo or checking on data to make sure that these two things hold or not.
- ④ Strong first stage could be tested, because we have data on Z and D , then we could just check whether the causal effect of Z on D is non-zero.
- ⑤ Monotonicity could be checked by plotting these variables to see the effects of IV is in the same direction on all individual units or not.

- (d) ① SUTVA: If one inmate knows that which therapist is "easier" on judging their mental health and more likely to put them into mental health courts, and this inmate tells others about this and they know which date and time is this therapist work, then the potential outcome for each inmates will be related to the treatment.
- ② Random Assignment: same story with the SUTVA, if inmates intentionally chose one therapist to make the judgement, then this couldn't be held.
- ③ Exclusion restriction: If the $\text{Cov}(Z, U) \neq 0$ or Z affects y through ways other than affect D ; this assumption couldn't hold.



④ Strong first stage: If we run a regression of Z on D , and the coefficient is not statistically significant, then this assumption couldn't hold, just like the classic example of "first 2 children of same gender make people more likely to have a third".

⑤ Monotonicity: For example, if the instrument, which is the therapist's choices of put inmates into mental health court is not in the same direction of treatment status; in this case, means that if therapist choose this inmate to go to mental health court, but then this inmate goes to normal courts, then this assumption didn't hold.

$$(e) \text{ Causal model: } Y_i = S_i + \varepsilon$$

$$\text{first step regression: } S_i = \Pi_{10} X + \Pi_{11} Z_i + \eta_{1i}$$

$$\text{reduced form: } Y_i = \Pi_{120} + \Pi_{121} Z_i + \eta_{2i}$$

where Z is the instrument of choices therapist made

① Y is the recidivism rate of each inmates

② X is some other unogenous variables affect the therapists choices

(4)

(f) The parameter that are estimated for S_i is basically the ratio of $\text{cov}(Z, Y)$ and $\text{cov}(Z, S)$. Given the model I had in (e), parameter of 2SLS is basically $\text{cov}(S, t)/\text{cov}(S)$. Since we couldn't estimate those other parameters in the real world, such as ATE and ATT, and since we don't have counterfactual states or even couldn't observe those, in ATT equation, we could only have values of $E[Y_0 | D=1]$ so we can not estimate it. But 2SLS estimators doesn't rely on the counterfactuals, so it can be estimated.

(g) $E[\hat{\beta}_{2SLS} - \beta] \approx \frac{\sigma_{v\eta}}{\sigma_\eta^2} F + 1$, where v is the residual in causal model and η is the residual in first stage equation.

We could test this with an F test on the joint significance of Z in the first stage.

Question 3